THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 3, March 1977
*Printed in U.S.A.*

# Pitch-Adaptive DPCM Coding of Speech With Two-Bit Quantization and Fixed Spectrum Prediction

By N. S. JAYANT

(Manuscript received June 9, 1976)

*This paper is concerned with the utilization of speech waveform periodicities in differential pulse code modulation (DPCM) coding with 2-bit adaptive quantization and time-invariant spectrum prediction. Our work is based on computer simulations of DPCM codes. We have studied pitch detectors based on autocorrelation and an average magnitude difference function (AMDF), and we have measured the benefits of predicting from a previous pitch period as functions of pitch-period-updating frequency and periodicity-indicating thresholds (for autocorrelation and the AMDF). We have compared several alternative methods of utilizing past quantized samples (in the present and previous pitch periods) for providing speech sample predictions. We find the following combination to be attractive for waveform coding at bit rates in the neighborhood of 16 kb/s: 2-bit adaptive quantization with a one-word (2-bit DPCM word) memory, pitch detection performed on unquantized speech (preferably with an AMDF criterion) and a prediction scheme that uses fixed three-tap (short-term) prediction for nonperiodic waveform segments, but switches to an appropriate one-tap (long-term) predictor upon the detection of strong periodicity. With four sample utterances, the latter procedure results in an average SNR (signal-to-noise ratio) gain of 3.75 dB over a non-pitch-adaptive encoder.*

## I. INTRODUCTION

An important subclass of speech waveform encoders is characterized by the use of adaptive quantization and predictive (DPCM) encoding.[1] Time-invariant spectrum predictors are simple to implement and robust in the context of coarse quantization. The benefits of adaptive prediction are, however, well recognized and documented,[2,3] and the greatest

achievements in bit-rate reduction bave in fact depended on the use of adaptive short-term (spectrum) prediction as well as adaptive long-term (pitch) prediction, as seen in the paper by Atal and Schroeder.[4]

This paper is concerned with the relatively less documented combination of *adaptive pitch prediction and nonadaptive spectrum prediction*. The study of this kind of prediction is motivated by the observation that speech waveforms abound in highly periodic segments and by the conjecture that the use of this periodicity may provide a prediction potential that is substantial enougb to obviate the need for adaptive short-term (spectrum) prediction. The attraction in this approach will evidently depend on the complexity of pitch detection itself. The pitch detectors used in this paper are based on autocorrelation and AMDF (average magnitude difference function) and are quite simple to implement; they are indeed much simpler than the mean-squared-error-minimizing pitch detector described in Ref. 4. Moreover, as discussed in Section IV, the success of pitch-adaptive DPCM does not depend critically on accurate pitch detection in the sense in which the term is used in formal speech research.[5]

A thesis by Trottier[6] considers the possibility of simplifying the Atal-Schroeder encoder.[4] Among other things, this thesis discusses simple pitch-detection algorithms, the criticality of a well-designed adaptive quantizer, and the inefficiency of approaches seeking to simplify adaptive spectrum prediction through the use of very few predictor taps, say two. An unpublished work of Grizmala[7] provides one of the first proposals for a simple pitch-adaptive DPCM that entirely avoids adaptive spectrum prediction. Grizmala discusses AMDF-based pitch detection and fixed three-tap spectrum prediction for nonperiodic waveform segments. More recently, Xydeas and Steele report an instance of a 6-dB SNR gain for a fixed-spectrum DPCM encoder arising from the utilization of waveform periodicities.[8] Finally the detection of periodicity based on autocorrelation and AMDF is documented in speech papers[5,9,10] as well as in coding literature.[11]

One of the contributions of the present paper is the demonstration that fixed-spectrum pitch-adaptive DPCM is useful in the context of a specific type of adaptive quantizer that has received considerable attention in recent coding work.[12,13] This paper also shows that AMDF-based pitch detection is slightly more effective than an autocorrelation-based procedure. The paper also demonstrates that, during periodic waveform segments, a simple one-tap predictor across the pitch period is more efficient than several multitap predictors involving many past samples in the present and previous pitch periods. Finally, the paper includes formal measurements of pitch prediction gain as a function of (*i*) pitch-period-update frequency, and of (*ii*) thresholds that the AMDF and correlation functions should exceed for a waveform segment to be

judged as periodic. Our results are all based on computer simulations of DPCM encoders.

The results of this paper are expected to be relevant to speech waveform coding at bit rates in the order of 16 kb/s. At this bit rate, the use of fixed spectrum prediction and adaptive quantization results typically in a quantization noise level that is quite easily perceived, while the sophistication of adaptive spectrum prediction is often unwarranted, because undesirable quantizer-predictor interactions begin showing up at around 16 kb/s in practical waveform coder designs.[14,15] Adaptive pitch prediction, on the other hand, appears to be a useful and robust sophistication at 16 kb/s. With this bit rate in mind, this paper will deal exclusively with two-bit quantizers for the DPCM coding of Nyquist-sampled (8-kHz) telephone-quality (200–3200 Hz) speech. Our numerical results refer to two female utterances, "The chairman cast three votes" and "The boy was mute about his task," and two male utterances "A lathe is a big tool," and "The boy was mute about his task." These utterances will henceforth be labeled F1, F2, M1, and M2.

The organization of the paper is as follows. Section II recommends a slowly adaptive quantizer with a one-word memory, and Section III proposes a three-tap spectrum predictor. Section IV discusses pitch detection by means of AMDF- and autocorrelation-type procedures, and points out how pitch analysis can be performed either on quantized speech or on the original unquantized speech. Section V compares different prediction algorithms for periodic segments, including the important example of an appropriate one-tap predictor. Section VI measures the gains of pitch-adaptive DPCM as a function of (*i*) the pitch-detection procedure, (*ii*) AMDF and autocorrelation thresholds used in hypothesizing periodicity, (*iii*) pitch-period-updating time, and (*iv*) prediction algorithms used for periodic waveform segments. Section VII summarizes performance figures for the four sample sentences and discusses results in the context of 16-kb/s waveform-coding.

## II. TWO-BIT ADAPTIVE QUANTIZER

Figure 1 shows a uniform four-level quantizer used for pitch-adaptive DPCM coding. The step-size $\Delta$ is adaptive. The adaptations are based on a one-word memory.[12,13] Specifically, the step-size is modified at every sampling instant by a multiplier that depends only on whether the magnitude of the previous quantizer output was $0.5\Delta_r$ or $1.5\Delta_r$. Respective step-size multipliers make $\Delta_{r+1} = E_1 \cdot \Delta_r$ or $E_2 \cdot \Delta_r$. In the context of quantizing prediction errors across a pitch period, we have found that the most useful adaptations were 'slow' adaptations of the form:[12]

$$E_1 = 0.95; E_2 = 1.10. \tag{1}$$

As discussed at length in Ref. 12, values of optimal step-size multipliers

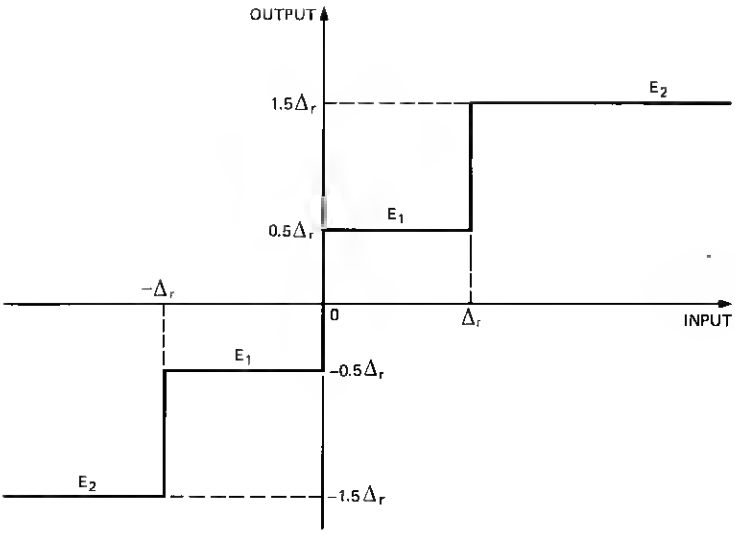


Fig. 1—A 2-bit adaptive quantizer.

reflect the nature of the input signal spectrum, and the stationarity of the input variance. The step-size adaptations were subject to maximum and minimum values that were appropriate for the given peak speech amplitude of ±1024:

$$\Delta_{MAX} = 192, \quad \Delta_{MIN} = 1.5. \tag{2}$$

Finally, nonuniform quantizers were not found to be very effective in pitch-adaptive DPCM using adaptive quantization. This had to do with the effect of DPCM predictions on the probability density function (PDF) at the quantizer input. The observation that nonuniform quantization is not very beneficial reflects the fact that predictions in DPCM cause a quantizer-input PDF that is more gaussian than the PDF of the original speech amplitudes. The latter, for example, can be modelled by a gamma-PDF for which nonuniform quantization is very useful.[2,3]

## III. TIME-INVARIANT SPECTRUM PREDICTION

A $T$-tap spectrum predictor is represented by

$$X_r = \sum_{s=1}^{T} a_s \cdot XQ_{r-s}, \tag{3}$$

where $X$ and $XQ$ refer to input and quantized speech samples.

In time-invariant (fixed) prediction, the coefficients $a$ are matched to the long-term spectrum of speech via the corresponding autocorrelation function, as described in Ref. 1.

Using a typical long-term spectrum characterization,[7] the following designs have been used for fixed one-tap and three-tap spectrum predictors:

$$a_1 = 0.85 \qquad \text{for } T = 1 \tag{4}$$

and

$$a_1 = 1.10; \quad a_2 = -0.28; \quad a_3 = -0.08 \qquad \text{for } T = 3. \tag{5}$$

These predictor coefficients are rounded values resulting from a spectrum model where the speech autocorrelations are 0.825, 0.562, and 0.308 for delays of one, two, and three 8-kHz samples, respectively. These autocorrelations are reported in Ref. 16 as the result of a study on a very large speech-sample base, and constitute slight revisions of very similar autocorrelations reported in Ref. 17.

In coding our speech waveforms, the three-tap predictor provided a typical SNR gain of nearly 1 dB over the one-tap predictor. Spectrum predictions in this paper will henceforth refer to a time-invariant three-tap design, as in eq. (5).

## IV. MEASUREMENT OF PITCH PERIOD

This section defines the AMDF- and autocorrelation-based pitch measurements used in our work, discusses the use of unquantized speech samples $X$ or quantized samples $XQ$ for the pitch analysis, and provides illustrations of pitch measurements. In general, pitch analysis will be based on a window $\mathcal{W}$ containing $W$ contiguous speech samples $Z$ ($Z = X$ or $XQ$). The sampling instant when a pitch period is measured is denoted by $r$, so that a current speech sample will be $Z_r$ ($X_r$ or $XQ_r$, as appropriate). The pitch period is denoted by $P$, and $P$ is assumed to have minimum and maximum values $P_{\text{MIN}}$ and $P_{\text{MAX}}$, respectively. $G_1$ and $G_2$ are thresholds that can be used to hypothesize waveform periodicity with varying degrees of confidence. $V$ is the pitch period updating time (see Section VI).

### 4.1 AMDF-based pitch measurement

Consider the average magnitude difference function

$$\text{AMDF}(p) = \text{AVERAGE}|Z_u - Z_{u-p}|;$$
$$p = P_{\text{MIN}}, P_{\text{MIN}} + 1, \cdots, P_{\text{MAX}}, \tag{6}$$

where the averaging is over all pairs $(u, u-p)$ such that both $Z_u$ and $Z_{u-p}$ are in $\mathcal{W}$.

The AMDF pitch detector estimates the pitch period $P$ to be

$$P = p_{\text{EST}}$$

$$\text{if } \text{AMDF}(p_{\text{EST}}) < \text{AMDF}(p) \tag{7}$$

for all $p$ in the range $(P_{\text{MIN}}, P_{\text{MAX}})$ with the exception of $p_{\text{EST}}$, and if

$$\text{AMDF}(p_{\text{EST}}) < G_1 \cdot \text{AVERAGE}(|Z_u|), \text{ for} \tag{8}$$

all $u$ in $\mathcal{W}$.

The value of $G_1$ is discussed in detail in Section VI. Typically, $G_1 = 0.5$. With Nyquist-sampled (8-kHz) speech and for a single pitch-analysis procedure that should cover the expected range of $p$ in both male and female speech, the following numbers seem appropriate:[5]

$$P_{\text{MIN}} = 16, \quad P_{\text{MAX}} = 160, \quad W = 256. \tag{9}$$

Notice that $P_{\text{MIN}}$ excludes the obvious minimum AMDF (0) at $p = 0$, and that the window length $W$ is well in excess of the maximum anticipated pitch period $P_{\text{MAX}}$. It turns out that this requirement ($W > P_{\text{MAX}}$) is quite important for efficient pitch prediction and waveform coding. The range of the pitch-period search ($16 < p < 160$) is wide enough to cause frequent problems with multiple peaks in the AMDF function, and multiples of the fundamental pitch period are often picked up as $P$. Fortunately, however, this kind of error in pitch tracking appears to be quite harmless as far as pitch-adaptive waveform codes are concerned: the need is for a sequence of waveform samples $\{XQ\}$ that provide good predictions of a current sequence $\{X\}$ in periodic segments, and it seems to be immaterial whether $\{X\}$ and $\{XQ\}$ are one pitch period apart or $n$ ($>1$) pitch periods apart.

#### *4.2 Autocorrelation-based pitch measurement*

Consider the autocorrelation function

$$C(p) = \text{AVERAGE}(\text{sgn } Z_u \cdot \text{sgn } Z_{u-p});$$
$$p = P_{\text{MIN}}, P_{\text{MIN}} + 1, \cdots, P_{\text{MAX}}, \tag{10}$$

where the averaging is over all pairs $(u,u-p)$ such that both $Z_u$ and $Z_{u-p}$ are in $\mathcal{W}$ and, furthermore, both $|Z_u|$ and $|Z_{u-p}|$ exceed an appropriate speech-clipping level

$$Z_{\text{CLIP}} = 0.64 \text{ MAX}(|Z|^1_{\text{MAX}}, |Z|^3_{\text{MAX}}), \tag{11}$$

where $|Z|^1_{\text{MAX}}$ is the maximum speech magnitude in the first one-third part of $\mathcal{W}$ and $|Z|^3_{\text{MAX}}$ is the maximum speech magnitude in the third one-third part of $\mathcal{W}$.

The autocorrelation-pitch detector estimates the pitch period $P$ to be

$$P = p_{\text{EST}} \tag{12}$$

if

$$C(p_{EST}) > C(p)$$

for all $p$ in the range of $(P_{MIN},P_{MAX})$ with the exception of $p_{EST}$, and if

$$C(p_{EST}) > G_2. \tag{13}$$

The role of $G_2$ is discussed at length in Section VI. Typically $G_2 = 0.2$. Appropriate values of $P_{MIN}$, $P_{MAX}$, and $W$ follow (9). The nonzero value of $P_{MIN}$ excludes the obvious maximum $C(0)$ at $p = 0$.

The center-clipping operation described by (11) is quite effective in mitigating spurious peaks in the $C(p)$ function, such as peaks representing a low first-formant frequency. Typically, autocorrelation pitch detectors work with speech that is low-pass filtered to, say, 900 Hz,[5] but such filtering was not used in our waveform coding program.

The pitch-measurement techniques based on (6) and (10)—especially the autocorrelation method (10)—are easier to implement than the mean-squared-error-minimizing pitch detector described in Ref. 4, which is based on computing the autocorrelation of $Z$ [this involves computing products of real numbers, instead of taking differences as in (6) or using one-bit numbers as in (10)]. The efficacies of AMDF- and autocorrelation-based pitch detectors have recently been calibrated in terms of the performance of several other pitch-tracking procedures.[5]

#### *4.3 Pitch analyses based on X and XQ*

Figure 2a demonstrates pitch analysis based on original, unquantized speech samples $X$. We see how the analysis window can be aligned so as to extend equally on either side of the current sample $X_r$ to be encoded

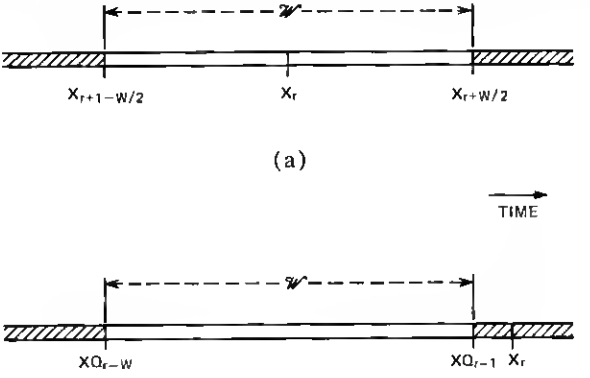


Fig. 2—Pitch analysis based on (a) unquantized speech $X$ and (b) quantized speech $XQ$.

Table I — Local and global minima/maxima in pitch-period search ($G_1 = 0.84$, $G_2 = 0.2$; speech sample: M2, analysis based on unquantized speech)

| $p$* | Minimization of Normalized AMDF | Maximization of Autocorrelation |
|---|---|---|
| 29 | — | 0.33 |
| 30 | 0.66 | 0.34 |
| 31 | — | 0.35 |
| 34 | 0.62 | — |
| 37 | — | 0.38 |
| 38 | — | 0.39 |
| 95 | — | 0.45 |
| 96 | 0.40 | 0.46 |

* Pitch-period estimate = 96 samples

(quantized); such alignment turns out to be quite critical for realizing the maximum potential of pitch-adaptive waveform codes.

Figure 2b shows the analysis of pitch based purely on past quantized samples $XQ_{r-s}$ ($s > 0$). Figures 2a and 2b apply equally to AMDF or autocorrelation analysis.

### 4.4 Illustrative measurements of pitch

Table I demonstrates examples of AMDF- and autocorrelation-based searches for the pitch period $P$. Entries in the table represent those local minima/maxima in the AMDF/$C$ functions, which were below/above all previous local minima/maxima in the search for $P$ ($16 < p < 160$). Also, only those minima/maxima that cross the $G_1/G_2$ thresholds, eqs. (8) and (13), are listed. For both the AMDF and $C$ functions, a global peak appears at the pitch period $P = 96$.

Table II provides a typical time plot of $P$ (number of 8-kHz samples) for four different pitch-tracking techniques. The analysis refers to a sample segment from the utterance F1. Notice the remarkable closeness of $X$-based contours in columns 1 and 3. Notice also that with $XQ$-based analyses, the AMDF function tends to preserve pitch information much better than the autocorrelation measurement.

## V. PREDICTION ALGORITHMS FOR PERIODIC WAVEFORMS

Figure 3 sketches a periodic waveform segment. $P$ is the 'pitch period', $X_r$ is a current waveform sample to be encoded, and $XQ$ denotes an already quantized sample in the present 'pitch period' or in an earlier 'very similar segment' of the periodic waveform.

Our prediction algorithms for periodic waveforms are linear, and they are of the general form

$$\hat{X}_r = \sum_{u=1}^{3} a_u \cdot XQ_{r-u} + \sum_{v=0}^{3} a_{P+v} \cdot XQ_{r-P-v}. \tag{14}$$

Table II — Pitch-period contours from four pitch-tracking techniques (speech sample: F1). Entries along columns are successive values of $P$ (number of 8-kHz samples)

| AMDF of $X$ | AMDF of $XQ$ | Autocorrelation of $X$ | Autocorrelation of $XQ$ |
|---|---|---|---|
| 2 | 2 | 2 | 19 |
| 39 | 39 | 2 | 19 |
| 78 | 39 | 78 | 19 |
| 39 | 39 | 39 | 19 |
| 39 | 39 | 39 | 39 |
| 39 | 39 | 39 | 38 |
| 39 | 39 | 39 | 39 |
| 39 | 2 | 39 | 41 |
| 43 | 2 | 43 | 44 |
| 40 | 2 | 40 | 35 |
| 41 | 42 | 41 | 25 |
| 132 | 132 | 132 | 2 |
| 134 | 134 | 134 | 2 |
| 135 | 134 | 135 | 2 |
| 57 | 135 | 57 | 2 |
| 78 | 80 | 78 | 48 |
| 157 | 157 | 157 | 50 |
| 35 | 35 | 2 | 19 |
| 2 | 2 | 2 | 19 |
| 2 | 2 | 2 | 19 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 18 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 18 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 31 |
| 35 | 2 | 35 | 34 |
| 35 | 2 | 35 | 34 |
| 35 | 35 | 35 | 35 |
| 36 | 35 | 36 | 36 |
| 36 | 35 | 36 | 36 |
| 36 | 36 | 36 | 36 |
| 37 | 36 | 37 | 37 |
| 37 | 37 | 37 | 37 |
| 37 | 37 | 37 | 37 |
| 37 | 37 | 37 | 37 |
| 37 | 37 | 37 | 37 |
| 37 | 37 | 37 | 37 |
| 37 | 37 | 37 | 37 |
| 75 | 37 | 75 | 37 |
| 75 | 75 | 37 | 37 |
| 37 | 75 | 37 | 37 |

We have considered many special cases of the general algorithm (14); Table III summarizes three interesting examples.

The seven-tap predictor attempts a clever combination of spectrum prediction [see (5) in Section III] and pitch prediction. This approach was proposed by Grizmala,[7] who in turn was simplifying a formal procedure of Atal and Schroeder.[4] The three-tap predictor is the simplest

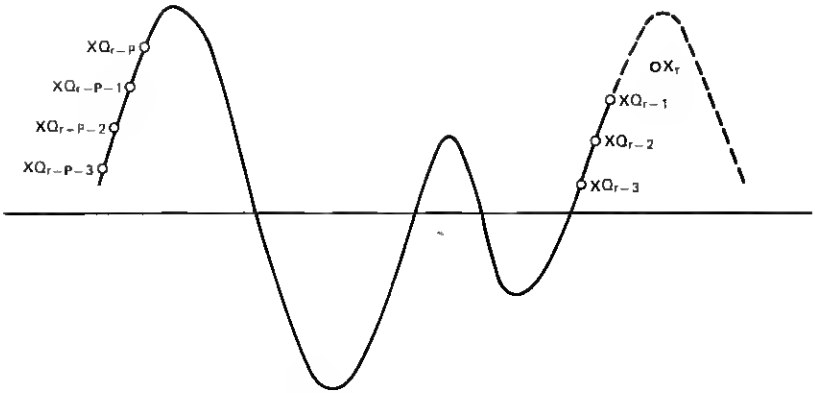


Fig. 3—Prediction algorithms for periodic waveforms.

nontrivial combination of the two types of prediction. It is suggested by a simple geometrical procedure of completing an idealized parallelogram with vertices at the topmost four dots in Fig. 3. Finally, the one-tap predictor is the simplest approach to pitch-adaptive coding and is suggested by the very strong correlations that are observed between $X_r$ and $X_{r-P}$ in highly periodic waveform segments.

## VI. DESIGN AND PERFORMANCE OF PITCH-ADAPTIVE DPCM CODER

Figure 4 provides a block diagram of the pitch-adaptive DPCM coder. It is different from conventional DPCM[1] in the inclusion of a special predictor for encoding the periodic segments of the input waveform. The spectrum predictor is formally defined by (5) and the pitch predictor by (14). The switching between the two predictors is controlled by the crossings of appropriate thresholds $G_1$ and $G_2$ (Section IV) by the AMDF or autocorrelation functions, respectively. The test for periodicity is done once every $V$ samples. If the waveform is decided to be "periodic" as a result of the test, the pitch period $P$ (coming out of the AMDF or autocorrelation measurement) is used in the predictive encoding of a current block of $V$ samples. (Both the binary "periodic/nonperiodic" decision and the pitch period, if any, are updated for the next block of $V$ samples.)

### *6.1 SNR, SNRV, and SNRSEG*

The design and utility of pitch-adaptive coders will be discussed using the following signal-to-noise ratio as a performance criterion

$$\text{SNR(dB)} = 10 \log_{10}\left[\sum_{r=1}^{N} X_r^2 \Big/ \sum_{r=1}^{N} (X_r - XQ_r)^2\right], \qquad (15)$$

Table III — Three prediction algorithms for periodic waveforms

| Name of Predictor | $a_1$ | $a_2$ | $a_3$ | $a_P$ | $a_{P+1}$ | $a_{P+2}$ | $a_{P+3}$ |
|---|---|---|---|---|---|---|---|
| AVERAGER | 0.5 | 0 | 0 | 0.5 | 0 | 0 | 0 |
| "7-Tap" | 1.1 | −0.28 | −0.08 | 1 | −1.1 | 0.28 | 0.08 |
| "3-Tap" | 1 | 0 | 0 | 1 | −1 | 0 | 0 |
| "1-Tap" | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

where $N$ is the total number of input samples.

In deference to the fact that the pitch-adaptive coding is performed in blocks of $V$ samples, we consider an additional measure of performance for the $S$th block

$$\text{SNRV}(S)(\text{dB}) = 10 \log_{10} \left[ \sum_{r=V(S-1)+1}^{V \cdot S} X_r^2 \Big/ \sum_{r=V(S-1)+1}^{V \cdot S} (X_r - XQ_r)^2 \right]. \quad (16)$$

The average value of SNRV over the total input signal duration (over $N/V$ input blocks) will be called the 'segment-signal-to-noise ratio' SNRSEG (Ref. 18)

$$\text{SNRSEG} = \frac{1}{N/V} \sum_{S=1}^{N/V} \text{SNRV}(S). \quad (17)$$

SNRV is an obvious indicator of local encoding quality; its average value SNRSEG reflects aspects of quantizer performance that do not

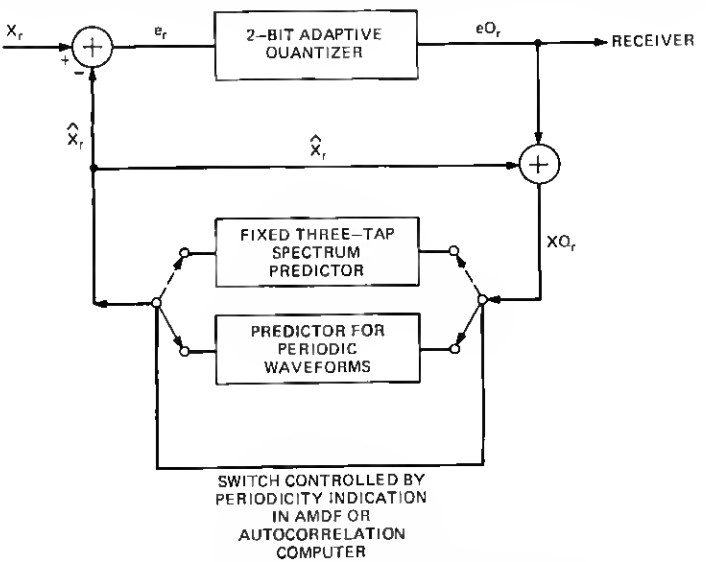


Fig. 4—Block diagram of pitch-adaptive coder.

Table IV — Comparison of prediction algorithms (utterance: F1; number of blocks: 134; block length $V$: 64; pitch-detector: based on unquantized speech and AMDF; $G_1 = 0.71$)

| Predictor | Averager | 7-Tap | 3-Tap | 1-Tap |
|---|---|---|---|---|
| SNR(dB) | 10.3 | 13.1 | 13.3 | 14.4 |
| SNRSEG(dB) | 15.0 | 16.5 | 16.8 | 16.8 |

always come out from the conventional SNR measure.[18] For example, the time variation of SNRV would provide an appropriate indication of the differential treatment of voiced and unvoiced waveform segments (this is seen in Fig. 5); also, occasional large samples of SNRV (associated with pitch-adaptive coding of highly periodic segments) would have a hetter chance of showing up in the final result if the performance measure is SNRSEG, rather than the conventional SNR.

### *6.2 Comparison of the prediction algorithms of Table III*

Tahle IV compares the performances of the four predictors in Tahle III for the DPCM encoding of a typical position of utterance F1. It is very interesting that the simplest of these predictors, the one-tap predictor, provides the hest encoding. In fact, the rest of this paper will uniformly assume an appropriate one-tap predictor for periodic segments.

### *6.3 Choice of decision thresholds $G_1$ and $G_2$*

Tahle V illustrates AMDF-hased coding as a function of the periodicity-decision threshold $G_1$ [see (8)]. A choice of $G_1 = 0.84$ appears to provide the hest comhination of SNR and SNRSEG. This value of $G_1$ corresponds to a 1.5-dB prediction gain [ratio of average magnitude of input $X$ to average magnitude of prediction error e (see Fig. 4)]. The value of $G_1 = 0.71$ (corresponding to a 3-dB prediction gain) provides a performance that is very close to the maximum. In fact, Grizmala[7] recommends the latter value of $G_1 = 0.71$.

Tahle VI shows corresponding results for autocorrelation-hased DPCM with $G_2$ as parameter. One notes a hroad optimum, with $G_2 = 0.2$ rep-

Table V — Effect of $G_1$ on AMDF-based pitch-adaptive DPCM (all parameters are the same as for Table I except that $G_1$ is now a variable)

| $G_1$ | 0 | 0.50 | 0.71 | 0.84 | 1.0 |
|---|---|---|---|---|---|
| SNR(dB) | 9.3 | 14.2 | 14.2 | 14.4 | 14.5 |
| SNRSEG(dB) | 12.5 | 15.2 | 16.6 | 16.8 | 14.9 |

Table VI — Effect of $G_2$ on autocorrelation-based pitch-adaptive DPCM (all parameters are the same as for Table I except that the pitch detection is now correlation-based)

| $G_2$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.6 |
|---|---|---|---|---|---|
| SNR(dB) | 13.6 | 13.8 | 13.6 | 13.3 | 10.3 |
| SNRSEG(dB) | 14.6 | 14.5 | 14.3 | 15.8 | 14.3 |

resenting a reasonable autocorrelation threshold for hypothesizing periodicity; it is interesting that an SNRSEG criterion would dictate $G_2 = 0.4$.

### *6.4 Comparison of pitch detectors: AMDF vs autocorrelation; X-analysis vs XQ-analysis*

Table VII compares, for optimal settings of $G_1$ and $G_2$, the encoding performances of AMDF- and autocorrelation-based pitch measurements. Notice the slight superiority of the AMDF approach, especially from an SNRSEG point of view. Notice also that pitch analyses based on X (Fig. 2a) are distinctly superior to those based on quantized speech $XQ$ (Fig. 2b). Finally, it is very significant that, in the case of $XQ$-based analyses, the value of SNRSEG is 3- to 5-dB higher than that of SNR. This indicates that even with $XQ$-based designs, many periodic segments get encoded very well in a short-term sense (leading frequently to very good SNRV values that tend to boost the average SNRV-value SNRSEG). The above observation has been confirmed in informal listening tests. These tests have also shown that the quantization noise in $XQ$-based AMDF-coding tends to be "whiter" than the noise obtaining with the other three pitch-detection schemes of Table VII.

### *6.5 Pitch-period update-time V*

Table VIII shows coder performance as a function of how frequently the periodicity test is made, and a possible pitch period recomputed.

Table VII — Comparison of four pitch detectors (all parameters are the same as for Table I, except that four pitch detectors are involved, and $G_1$ and $G_2$ are optimized for each case)

| Type of Pitch Analysis | AMDF | | Autocorrelation | |
|---|---|---|---|---|
| Basis of the analysis | $X$ | $XQ$ | $X$ | $XQ$ |
| SNR-optimizing G-values ($G_1$ for AMDF, $G_2$ for correlation) | 0.84 | 0.84 | 0.20 | 0.30 |
| SNR(dB) | 14.4 | 10.0 | 13.8 | 10.1 |
| SNRSEG(dB) | 16.8 | 15.0 | 14.5 | 13.2 |

Table VIII — Dependence of performance on update time $V$; entries are SNR values in dB (female utterance: F1; number of blocks: 134; male utterance: M1; number of blocks: 134; pitch detector: based on unquantized speech and AMDF; $G_1 = 0.71$)

| $V$ | 32 | 64 | 128 | 192 |
|---|---|---|---|---|
| Male | — | 12.1 | 11.4 | 9.8 |
| Female | 15.1 | 14.4 | 12.8 | — |

Recall that the update time assumed in Tables IV through VII was $V = 64$ samples (8 ms). Previous researchers[4–7] have usually recommended $V$-values like 40 or 50.

## VII. SUMMARY AND CONCLUSIONS

Table IX compares, for the complete utterances F1, F2, M1, and M2, the performance of pitch-adaptive DPCM coding with that of DPCM with a fixed three-tap spectrum predictor. Note that both of these coders use adaptive quantization. The conventional encoder uses a fixed spectrum predictor while the pitch-adaptive encoder includes a second adaptive one-tap predictor, which is switched in whenever an AMDF analysis on $X$ suggests sufficient periodicity ($G_1 = 0.84$).

We note that there exists across the four sample sentences an average 3.8-dB SNR gain with pitch-adaptive coding. The better performance with female speech is not surprising, since for a given duration of a voiced speech utterance, the high-pitched female utterances have a greater number of pitch periods.

Figure 5 provides a typical time-plot of pitch period P and local signal-to-noise-ratio SNRV in pitch-adaptive coding. The example refers to a segment from F2. A pitch-period of zero in Fig. 5 indicates absence of periodicity. Notice the low values of SNRV for these nonperiodic blocks. Also, notice the cluster of three values of $P \simeq 133$. These three estimates are obviously three times a true pitch period $\simeq 44$.

As mentioned earlier, the work in this paper was motivated by the desire to improve waveform encoder performance at bit rates in the order

Table IX — Summary of DPCM encoder performance

| Sample Utterance | Median Pitch (Number of 8-kHz Samples) | Number of Speech Blocks ($V = 64$) | DPCM With no Pitch Tracking SNR(dB) | Pitch-Adaptive DPCM SNR(dB) |
|---|---|---|---|---|
| F1 | 36 | 240 | 10.0 | 15.0 |
| F2 | 40 | 288 | 14.0 | 18.0 |
| M1 | 90 | 192 | 11.0 | 13.5 |
| M2 | 92 | 245 | 11.0 | 14.5 |

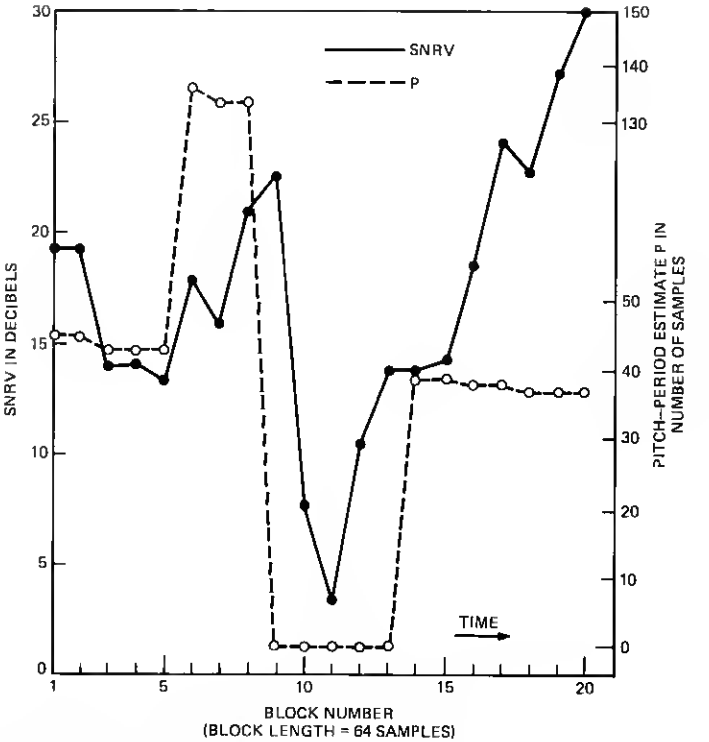


Fig. 5—Typical time variations of pitch period and local signal-to-noise ratio SNRV. (Data refers to a segment from utterance F2)

of 16 kb/s. The 2-bit pitch-adaptive coders discussed need 16 kb/s to transmit prediction-error information; and if pitch-analysis is to be performed on uncoded speech, the transmission of this information to a receiver will entail an additional channel capacity of about 1 kb/s. This assumes that pitch-period samples are coded with 7-bit accuracy and updated (and transmitted once, say, every 56 samples (8 kHz × 7 bits/56 = 1 kb/s). Alternatively, the coder can be used on a 16-kb/s channel if the sampling rate can be restricted to 15 kb/s/2 bits = 7.5 kHz.

## VIII ACKNOWLEDGMENT

The author wisbes to thank Mrs. I. Sondhi and Dr. Ing. P. Noll for their help with computer programs for pitch-adaptive coders.

## REFERENCES

1. N. S. Jayant, "Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers," Proc. IEEE, *62*, (May 1974), pp. 611–632.
2. P. Noll, "Non-adaptive and adaptive DPCM of speech signals," Polytech. Tijdschr. Ed. Elektrotech./Electron. (The Netherlands), No. 19, 1972.

3. P. Noll, "A Comparative Study of Various Quantization Schemes for Speech Encoding," B.S.T.J., *54*, No. 9 (November 1975), pp. 1597–1614.
4. B. S. Atal and M. R. Schroeder, "Adaptive Predictive Coding of Speech Signals," B.S.T.J., *49*, No. 8 (October 1970), pp. 1973–1986.
5. L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McConegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," IEEE Trans. Audio and Speech Signal Processing, *ASSP-24* (October 1976), pp. 399–418.
6. L. I. Trottier, "An Investigation of Digital Vocoders," Masters Thesis, Department of Electrical Engineering, McGill University, Montreal, Quebec, January 1973.
7. F. Grizmala, "Application of Linear Predictive Coding to Long Haul Facilities—Results of a Simulation Study," unpublished work, January 1972.
8. C. S. Xydeas and R. Steele, "Pitch Synchronous 1st-Order Linear D.P.C.M. System," Electron. Lett., *12* (19 February 1976), pp. 93–95.
9. M. M. Sondhi, "New Methods of Pitch Extraction," IEEE Trans. Audio Electroacoust., *16* (June 1968), pp. 262–266.
10. H. Fujisaki, "Pitch Measurement Using Autocorrelation Techniques," J. Accoust. Soc. Amer., *28* (1956), p. 1518(A).
11. M. J. Ross et al., "Absolute Magnitude Difference Function Pitch Extractor," IEEE Trans. Acoust. Speech Signal Process. *22* (October 1974), pp. 353–362.
12. N. S. Jayant, "Adaptive Quantization With a One-Word Memory," B.S.T.J., *52*, No. 7 (September 1973), pp. 1119–1144.
13. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J., *52*, No. 7 (September 1973), pp. 1105–1118.
14. S. U. H. Qureshi and G. D. Forney, "A 9.6/16 KBPS Speech Digitizer," Proc. IEEE Inter. Conf. Commun., San Francisco, June 1975, pp. 30-31 to 30-36.
15. D. L. Cohn and J. L. Melsa, "The Residual Encoder—An Improved ADPCM System for Speech Digitization," pp. 30-26 to 30-31 of Proceedings in Ref. 14.
16. R. P. Crane and E. T. Hedin, Jr., Bell Laboratories, unpublished work.
17. R. A. McDonald, "Signal-to-Noise and Idle Channel Performance of DPCM systems—Particular Application to Voice Signals," B.S.T.J., *45*, No. 7 (September 1966), pp. 1123–1151.
18. P. Noll, "Adaptive Quantizing in Speech Coding Systems," Int. Zurich Seminar on Digital Communication (IEEE), March 1974, pp. B3.1–B3.6.